

## Criteria as a Prelude for Guiding Taxonomy Evaluation

Daniel Szopinski  
Paderborn University  
[daniel.szopinski@  
wiwi.uni-paderborn.de](mailto:daniel.szopinski@wiwi.uni-paderborn.de)

Thorsten Schoormann  
University of Hildesheim  
[thorsten.schoormann@  
uni-hildesheim.de](mailto:thorsten.schoormann@uni-hildesheim.de)

Dennis Kundisch  
Paderborn University  
[dennis.kundisch@  
wiwi.uni-paderborn.de](mailto:dennis.kundisch@wiwi.uni-paderborn.de)

### Abstract

*Taxonomies are design science artifacts used by researchers and practitioners to describe and classify existing or future objects of a domain. As such, they constitute a necessary foundation for theory building. Yet despite the great interest in taxonomies, there is virtually no guidance on how to rigorously evaluate them. Based on a literature review and a sample of 446 articles, this study explores the criteria currently employed in taxonomy evaluations. Surprisingly, we find that only a minority of taxonomy building projects actually evaluate their taxonomies and that there is no consistency across the multiplicity of criteria used. Our study provides a structured overview of the taxonomy evaluation criteria used by IS researchers and proposes a set of potential guidelines to support future evaluations. The purposeful and rigorous taxonomy evaluation our study advances contributes to DSR by bridging the gap between generic evaluation criteria and concrete taxonomy evaluation criteria.*

### 1. Introduction

To enable information systems (IS) to develop its full potential, design science research (DSR) provides the concepts and methods required to explore and understand IS artifacts [1]. Representing one of the fundamental types of design science artifacts, taxonomies can be used to describe and classify existing or future objects of a domain. They also allow to differentiate objects by revealing similar and different characteristics, which is essential to understanding a domain of interest [2]. In this way, taxonomies serve to structure and organize a body of knowledge and facilitate the study of relationships among concepts, “with all the potential advantages that bring for the advancement of the field” [3, p. 85], and to hypothesize about these relationships. Enabling researchers to describe and analyze a domain is an inevitable aspect of theory building, for which taxonomies are well suited [2, 4, 5].

In striving for rigor, design science extensively evaluates artifacts [6, 7]. Like any other artifact in DSR [1], taxonomies too have to be evaluated [2]. Although as structure-giving artifacts they are important for the exploration of new research fields in IS (e.g., [8, 9]) and for developing novel software artifacts (e.g., [10]), taxonomies are rarely evaluated, nor is there any consistency in terms of the criteria used for evaluation. Nickerson et al. [2] have proposed guidance for taxonomy building in form of a development method, which is widely accepted in IS and has been applied to contexts as diverse as the internet-of-things [11], FinTechs [12], and crowdfunding [13]. However, we found that there is hardly any methodological guidance for taxonomy evaluation, and none whatsoever for evaluation criteria.

The evaluation of taxonomies responds to both the general call for evaluation in DSR (e.g., [14–16]), and to more taxonomy-specific calls for evaluation (e.g., [2, 17]). This study seeks to shed light on the question of *which criteria researchers currently employ to evaluate taxonomies*, and aims to inform future taxonomy endeavors. Based on a systematic analysis of existing literature, we make two main contributions: First, we provide an overview of the evaluation criteria researchers currently use to evaluate their taxonomies (i.e., artifact-specific evaluation criteria). Second, we make six guiding recommendations on the criteria which taxonomy developers should consider and how they can be applied meaningfully as part of taxonomy evaluations. We thereby extend the taxonomy development method proposed by Nickerson et al. [2] and enable researchers to undertake a more informed and purposeful evaluation. Our ultimate aim is to support researchers with “observ[ing] and measur[ing] how well the artifact [*in this case, taxonomies*] supports a solution to the problem [*in this case that of structuring a domain of interest*] [which] requires knowledge of relevant metrics [*here, taxonomy evaluation criteria*]” [18, p. 56].

## 2. Research background

### 2.1. Taxonomies in Information Systems

The ability to classify objects is of great importance to the organization of knowledge [19] and the analysis of complex domains [2]. Without it, each object has to be perceived as unique, and thus, researchers would be overwhelmed by the sheer diversity of objects [20]. For the classification of objects, schemes can be employed to, for instance, facilitate the structuring of concepts and their relationships to each other, or characterize similarities and differences between domains [2, 19, 21]. In IS research, taxonomies are seen as the prevailing form of classification schemes [2, 22].

In addition to their descriptive and classificatory purpose, a taxonomy can also serve as a *theory for analyzing*, the most basic type of theory (so-called taxonomic theory [5]) and the foundation for more advanced theories [4]. Similarly, as taxonomies attempt to describe the objects that play a role in certain phenomena [23], improving our understanding of a domain via taxonomies can also lead to basic theory-building as “a robust taxonomy can [...] be used to perform ex post theory building” [24].

As taxonomies are design artifacts, applying rigor to both their building and their evaluation has to be a prerequisite. Concerning taxonomy building, Nickerson et al. [2] propose an iterative, seven-step method which combines inductive and deductive building. This is, the method is able to develop a structure for understanding a domain of interest, deducing characteristics and dimensions (groups of characteristics) from conceptual foundations (e.g., previous research about a domain) as well as from empirical data (e.g., real-world objects from a domain). Concerning evaluation, the method comprises some guidance that advises to differentiate between objective and subjective ending conditions. Subjective ending conditions seek to ensure high quality while building taxonomies, and objective ending conditions determine when to stop the taxonomy development. It is unclear to what extent the ending conditions can also be used (or already are used) for the evaluation of completed taxonomies. Nickerson et al. [2] themselves “leave this as an area for future research” (p. 2).

As a starting point, researchers can draw on the rich body of DSR literature. Given that the “[criteria] for evaluation of IS designs and artifacts are too many to enumerate” [16, p. 84], we resisted the temptation to provide an exhaustive review of all generic evaluation criteria in DSR. Although there is a variety of general guidelines and criteria available in DSR (see [6-7], [25], [27]), there are only very few taxonomy-specific

evaluation guidelines, which is problematic because it makes it more difficult for researchers to make informed decisions with regard to the selection and suitability of evaluation criteria.

### 2.2. Evaluation criteria in Design Science

As part of conducting rigorous research on the development and use of artifacts, DSR emphasizes the role played by evaluation to check and confirm the robustness of research results (e.g., [6, 7, 14–16, 25, 27–28]). Suitable evaluation criteria should be chosen based on the nature of the artifact [18, 25]. Evaluations in DSR involve assessing how well an artifact achieves its expected utility, quality, and efficiency, as well as what makes it work. Furthermore, it may also involve comparing the performance of one artifact to that of another [16]. The *Framework for Evaluation in Design Science (FEDS)* offers four evaluation strategies accommodate the different nature of the risk that characterizes each type of artifact: that researchers can use to evaluate their artifacts [16]: (1) Low social or technical risks (*quick & simple*, e.g., for a taxonomy on hotel booking apps, [67]), (2) social or user-oriented risks (*human risk & effectiveness*, e.g., for a taxonomy on virtual reality in healthcare, [10]), (3) technical risks (*technical risk & efficiency*, e.g., for a taxonomy on smart things, [11]), and (4) without any social or human risks (*purely technical*, e.g., for a taxonomy on technical characteristics of block-chain systems, [26]). Thus, a taxonomy’s evaluation strategy is not only appropriate to the target group or purpose it serves, but also to the objects it intends to capture. To implement these abstract evaluation strategies, previous literature already advocates various context-independent (i.e., artifact-independent) methods and criteria. In an earlier study, we systematically analyzed methods that researchers employ to evaluate taxonomies [29] and concluded that there is no generally accepted consensus on the application of these methods with regard to the choice of suitable criteria for taxonomy evaluation. The selection of criteria for taxonomy evaluation is often ad-hoc and there is no prescriptive knowledge about which criteria are conducive to taxonomy evaluation. Therefore, this study focuses on the identification and analysis of evaluation criteria that reflect the specific characteristics of taxonomies.

## 3. Method

Pursuing the goal of a comprehensive overview of evaluation criteria which researchers currently employ to evaluate taxonomies, we obtained 446 articles in four separate steps (see Figure 1): (Step 1) A citation

analysis of the widely used method for taxonomy development in IS, the method by Nickerson et al. [2], (Step 2) a keyword search in the *AIS Senior Scholars' Basket of Journals*, (Step 3) a keyword search in the AIS eLibrary for articles published in the proceedings of the *ICIS* and *ECIS*, and (Step 4) a keyword search in the SpringerLink library (since 2010) for articles published in the proceedings of the *DESRIST*.

**Figure 1. Identification of relevant articles**

Phase 1: Identification		(Step 1) Citation analysis of Nickerson et al. [2]	(Step 2) Keyword search in AIS Senior Scholars' Basket	(Step 3) Keyword search in proceedings of ICIS/ECIS	(Step 4) Keyword search in proceedings of DESRIST	Total number of articles identified
	Total number of articles identified	252	70	73	51	446
	After removing non-English articles	237	70	71	51	431
	After removing neither journal nor conference	199	68	71	51	389
	After removing duplicates	191	63	52	50	356
Phase 2: Analysis	Articles that build a taxonomy	126	47	23	5	201
	Articles that evaluate a taxonomy	45	12	4	2	63
	Articles that report evaluation criteria	41	9	3	1	54

For the citation analysis of the seminal article by Nickerson et al. [2], we used the citation-indexing service by Google Scholar (Step 1). We sought to identify relevant articles of high quality by including articles that cite [2], that are written in English, and that are published in peer-reviewed journals or conference proceedings. Ending in October 2018, the citation analysis revealed 252 citations. We excluded those articles that do not match the aforementioned criteria and obtained 191 articles.

To identify additional articles, we performed a keyword-based search with “*taxonomy*” for title, abstract, and keywords *up until and including 10/2018* (timespan) in the *AIS Senior Scholars' Basket of Journals* (Step 2), in the proceedings of the *ICIS*, *ECIS* (Step 3), and *DESRIST* (Step 4). In total, we found 194 articles. We removed 24 duplicates and excluded five articles that are neither journal nor conference articles.

After creating our collated literature, the first and second author independently read each article and decided (i.e., coded) whether an article (1) is concerned with the building of a taxonomy (yes or no), (2) is concerned with the evaluation of a taxonomy (yes or

no) and (3) reports evaluation criteria (yes or no, if yes, which evaluation criteria). We identified 356 unique articles that match our initial search criteria. In 201 of these, researchers develop a taxonomy, in 63 articles they evaluate a taxonomy, and in 54 articles they report on the evaluation criteria. An article could appear in more than one of these three groups. For example, researchers may build and evaluate a taxonomy in an article. We validated the coding of the articles by calculating the interrater reliability. The following percentages of agreement were obtained: *taxonomy development* (98 %) (researchers build a taxonomy), *taxonomy evaluation* (97 %) (researchers evaluate a taxonomy), and *taxonomy evaluation criteria* (92%) (researchers name evaluation criteria). Reassuringly, the interrater agreement between the authors for the selected articles with regard to *taxonomy development* and *taxonomy evaluation* is very high throughout, with values for Cohen’s Kappa and Krippendorff’s Alpha between 0.89 and 0.96. With regard to *taxonomy evaluation criteria* values are lower but for the explorative character of the coding acceptable (0.62). In the cases of disagreement, the authors discussed their opinions to come to a joint verdict.

## 4. Taxonomy evaluation criteria

From the sample of 446 articles, we identified 54 articles that report on taxonomy evaluation criteria. Across the 54 articles, we identified 43 different evaluation criteria. In Table 1 we show which criteria were used, by, which article, and how often.

### 4.1. Objective and subjective ending conditions as taxonomy evaluation criteria

From our sample, researchers evaluate a taxonomy in 63 articles, and of these, 38 apply the taxonomy development method by Nickerson et al. [2]. Expectedly, these 38 articles should follow the objective and subjective ending conditions by [2] to indicate when to stop with the taxonomy building (i.e., objective conditions) and to verify the usefulness of the resulting taxonomy (i.e., subjective conditions).

**Objective ending conditions.** Referring to the eight objective conditions, 16 out of 38 articles explicitly state that they apply the ending conditions and do this without giving any further detail (e.g., [40]). An additional six articles offer details on how the ending conditions have been applied (i.e., describing the conditions in own words or adapting the descriptions from [2], e.g., [57, 64]). Another two justify when the ending conditions are perceived as fulfilled in the taxonomies’ particular context (e.g., “all repair processes have been examined [./] no repair

process will be combined or divided [./] no characteristic will be further added [./] no characteristic will be combined or divided [./] each characteristic will be unique”, [44, p. 421]). When analyzing the usage of the objective ending conditions, one observation stood out: There are articles which discuss the exclusion of some ending conditions. For instance, researchers explicitly exclude the ending condition ‘all objects or a representative sample of objects have been examined’, as the sample of objects employed in that particular research context is neither comprehensive nor representative [69]. In another example, the ending condition ‘at least one object is classified under every characteristic’ had been excluded because the characteristics of the taxonomy are conceptually derived (e.g., from latest research), but, not yet implemented by real-life objects [52].

**Subjective ending conditions.** With regard to the subjective ending conditions (i.e., concise, robust, comprehensive, extendible, and explanatory), four out of 38 articles merely cite the conditions (e.g., [70]). A further 20 articles state all or a subset of the subjective conditions, and another ten elaborate why and how evaluation criteria are fulfilled in the taxonomies’ particular context. By analyzing the application of the subjective ending conditions, we made the following observations: First, ten articles apply a subset of subjective conditions (e.g., [34]) or customize the set by, for example, adding, adjusting, and/or eliminating conditions (e.g., robust, concise, unambiguous, comprehensive, [33]). Second, ten articles provide details about why the evaluation criteria have been met by outlining their application (e.g., “*concise* enough to be easily applied [./] *robust* differentiation of distinct system archetypes [./] *comprehensive* in that it has allowed us to classify every crowdsourcing system in our samples [./] future work could certainly *extend* the

typology [./] current version has sufficient *explanatory* power with respect to the essential crowdsourcing mechanisms”, [36, p. 7]). A further three articles aim to measure the fulfillment of the criteria including, for example, by verifying the taxonomy’s robustness. This requires analyzing a taxonomy’s usage over time, which can be done in longitudinal studies in which a sample of objects is analyzed across several time periods [17]. Another article measures the taxonomy’s conciseness by determining its simplicity [25] in terms of, for example, the number of dimensions and characteristics. Third, some articles employ the subjective ending conditions, for example, to elaborate on the quality of other research outcomes (e.g., [63]), but without applying Nickerson et al.’s [2] method.

## 4.2. Usefulness as taxonomy evaluation criteria

Following the design science paradigm of building and evaluating artifacts, it is generally prescribed that the development of new artifacts requires determining its *usefulness*. Typically, *usefulness* is defined as “the degree to which the artifact positively impacts the task performance of individuals” [25, p. 266] and indicates the extent to which the design of an artifact is capable of usefully supporting stakeholders with their intended purpose of using the artifact [14]. There are hardly any guidelines, however, on how to evaluate the *usefulness* of taxonomies. The most frequently used evaluation criteria are those [2] propose as part of their taxonomy building method and which are supposed to “be applied before putting a taxonomy into use” [2, p. 342]. These criteria (i.e., the five subjective ending conditions) particularly aim to ensure the high quality of taxonomies during the process of taxonomy building.

However, some authors state that the *usefulness* of taxonomies can only be evaluated by observing a

**Table 1. Overview of identified articles**

Taxonomy evaluation criteria	References using taxonomy evaluation criteria	Appearance
<i>Usefulness</i>	[10–12, 17, 25, 30–58]	34
<i>Comprehensiveness*</i>	[11, 12, 25, 30–45, 59–64]	25
<i>Applicability</i>	[4, 11, 12, 24, 30–34, 43, 46–51, 59, 60, 65–70]	24
<i>Robustness*</i>	[11, 12, 17, 30–39, 43, 59–64, 71]	21
<i>Conciseness*</i>	[11, 12, 25, 30, 32, 34–44, 59–63]	21
<i>Extensibility*</i>	[11, 12, 25, 30–32, 34–41, 43, 44, 59–63]	21
<i>Explanatory*</i>	[11, 12, 30, 32, 34–41, 50, 59–63]	18
<i>Understandability</i> [4, 30, 31, 35, 50, 52, 56, 72], <i>collectively exhaustive</i> [4, 26, 35, 40, 43, 61, 71], <i>completeness</i> [25, 32, 34, 35, 53, 58, 73], <i>mutually exclusiveness</i> [26, 35, 40, 43, 61, 71], <i>distinctiveness</i> [26, 33, 43, 45, 46], <i>reliability</i> [11, 45, 54, 74], <i>utility</i> [10, 49, 61, 65], <i>efficiency</i> [10, 41, 51], <i>stability</i> [24, 57], <i>uniqueness</i> [39, 61], <i>repeatability</i> [38, 39], <i>sufficiency</i> [41, 70], <i>construct validity</i> [30, 75], <i>consistency</i> [45, 73], <i>parsimoniousness</i> [42, 45], <i>effectiveness</i> [17, 31]		2-8
<i>Exhaustiveness</i> [57], <i>adequateness</i> [35], <i>compatibleness with theories</i> [35], <i>purposefulness</i> [33], <i>unambiguousness</i> [33], <i>usability</i> [40], <i>inclusiveness</i> [40], <i>feasibility</i> [17], <i>descriptiveness</i> [44], <i>versatileness</i> [44], <i>sufficiently detailedness</i> [62], <i>generalizability</i> [54], <i>appropriate wording</i> [31], <i>relevance</i> [31], <i>real-world fidelity</i> [72], <i>face validity</i> [49], <i>modifiability</i> [25], <i>simplicity</i> [25], <i>suitability</i> [58], <i>no unnecessary categories</i> [4]		1

\*Subjective ending condition from Nickerson et al. [2]

taxonomy's application over time, and after completion (e.g., [36]), and would require additional evaluation criteria. This stipulation is also in line with design science literature, which advocates that a rigorous research process should strictly separate building from evaluation, especially when it comes to the iterative development processes of artifacts [7, 18, 28]. The evaluation criterion of *usefulness* is not only important for the time period chosen for the evaluation, but also for how it is defined. DSR literature critically discusses whether *usefulness* is an appropriate construct because it tends to focus on short- and medium-term horizons [28]. In view of the fact that taxonomies are often built to structure emerging concepts (e.g., block-chain), as part of theory development [4], one can assume that a rather long-term understanding of *usefulness* is more conducive to the evaluation of taxonomies. In our sample, *usefulness* is the most frequently cited evaluation criteria for taxonomies, used in 34 articles.

We observed various ways in which *usefulness* is operationalized in our sample. Whereas 17 articles merely name the term “useful” or “usefulness”, 17 provide more detail by explicitly stating target groups and/or intended activities. Regarding the target groups, the authors explain for *whom* a taxonomy is supposed to be useful (e.g., abstract groups such as researchers interested in developing new artifacts or theories, practitioners dealing with software engineering [40, 50], or specific groups such as analysts, investors, and standard setters, [38]). Regarding activities, *what* a taxonomy is supposed to be useful *for* (e.g., differentiation of objects [46], classification of objects [11, 12, 47], identification of objects [48], evaluation of objects [41], analysis of a domain [40], decision making [63], informing theory building [54], and understanding domains [50]). At the start of any taxonomy building, researchers should define both their target groups and their intended activities, and return to them again later as part of the evaluation, after the taxonomy building has been completed.

#### 4.3. Applicability as taxonomy evaluation criteria

The evaluation criterion *applicability* features in numerous taxonomy evaluations. *Applicability* refers to “evaluations by practice of the theories, models, frameworks, processes, technical artifacts, or other theoretically based IS artifact” [27, p. 2], indicating whether an artifact is applicable in practice [15]. Only very few guidelines are available to evaluate taxonomies with regard to their applicability. In our sample, 24 out of 54 articles demonstrate the *applicability* of their taxonomies by, for example, comparing the characteristics of real-life objects with

those of the taxonomy (e.g., already implemented software products, [65]). Others utilize illustrative scenarios (e.g., classify three assistance systems, [69]) and articles from related literature (e.g., classify a sample of journal articles, [4]) to demonstrate the applicability of the taxonomies. Similarly to usefulness, we observed various ways in which *applicability* has been operationalized in our sample. While a majority of 18 articles (of the 24) merely state that they verify the *applicability* of a taxonomy, six articles detail the meaning of *applicability* in their respective context. This subset comprises articles that are referring to, for example, a certain target group (e.g., applicable for practitioners related to risk management, [53]), and to the task for which the *applicability* is evaluated, such as the specification (e.g., to define Delphi studies, [33]), comparison (e.g., to compare platforms, [17]), or classifications of objects (e.g., to classify journal articles, [4]).

## 5. Discussion

As part of our study, we identify 202 articles that have built a taxonomy, noting that only 63 articles have evaluated their taxonomy after completion. We also found a plethora of very different evaluation criteria in use, across the sample (see Figure 2). As a result of our analysis, three main observations emerge. First, concerning the frequency of evaluation criteria, half of them appear only once in the sample – and hence are of little interest to the purpose of this study – and only a few are frequently used and can be said to dominate. Interestingly, among these criteria are some which find broader application in general DSR, such as *usability* [7] and *face validity* [14]. Second, as most studies follow the taxonomy development method by Nickerson et al. [2], the five subjective ending conditions of this method are often employed not only to determine the taxonomy building but also to evaluate the taxonomy after completion. However, two particular evaluation criteria, namely *usefulness* and *applicability*, stand out as the most frequently used. Third, the great heterogeneity of evaluation criteria – evidenced by 43 different criteria identified in our sample of 54 articles – underlines the need for common ground to be agreed upon by the research community on the criteria to be used to evaluate taxonomies.

We organize the discussion of our study's findings along six guidelines. By providing the latter we extend the support available to researchers who currently receive methodological support only for building, but not for evaluating their taxonomies. The ending conditions by [2] form the starting point of our discussion. With our analysis, we are revealing and summarizing the knowledge about taxonomy

evaluation criteria that is distributed across many domains and articles. The guidelines are intended for researchers who build and/or evaluate taxonomies. Guideline 1 can be used both while building and evaluating taxonomies, guidelines 2 and 3 for taxonomy building (i.e., ex ante evaluation), and guidelines 4 to 6 for ex post evaluation, i.e. once taxonomy building has been completed.

**Guideline 1:** *Scope your taxonomy evaluation.* This refers to defining what is to be evaluated (e.g., the entire taxonomy, the dimensions/characteristics of the taxonomy, and/or the descriptions of the dimensions/characteristics) and for whom (e.g., researchers with interest in a specific phenomenon, method or community and/or practitioners with interest in taxonomies to inform their decision-making for choosing among or developing software artifacts). For example, [77] scope their taxonomy evaluation by describing the specific target group of users (here, practitioners and academics) of their taxonomy and the tasks users are intended to perform by means of their taxonomy (here, analysis and selection of objects).

**Guideline 2:** *Justify the selection of the objective ending conditions and its degree of fulfillment when applying objective ending conditions.* This particularly addresses the question of whether to apply all or only a subset of ending conditions. It may make more sense to ignore some of the objective ending conditions, for instance, when the taxonomy is derived conceptually (i.e., from literature) and some of the characteristics suggested by the literature are not yet implemented (and thus cannot be observed) in real-world objects. Clearly stating which objective ending conditions have been applied, and why, facilitates the traceability of the evaluation and hence, future research based on that taxonomy. For example, [11] explicitly exclude one of the objective ending conditions (mutually exclusive) and justify this with a more readable and less complex taxonomy. In addition to this example, [44] discuss the

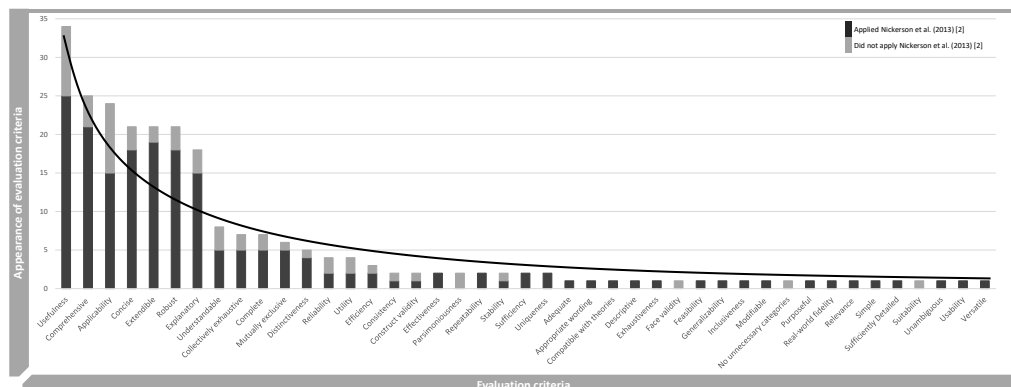
degree of fulfillment of the objective ending conditions in the context of the taxonomy's domain.

**Guideline 3:** *Contextualize the subjective ending conditions and their degree of fulfillment.* After completing taxonomy building, you should evaluate your taxonomy by employing the five subjective ending conditions by [2] and, provide information on how the evaluation criteria are fulfilled in the specific context of the taxonomy. Suggested criteria could include: being concise (e.g., taxonomy has seven plus/minus two dimensions), robust (e.g., when it can differentiate among objects), comprehensive (e.g., when all relevant objects can be described), extendible (e.g., when new dimensions or characteristics can be added), and/or explanatory (e.g., when it can explain objects). For example, [47] critically discuss the fulfillment of each of the subjective ending conditions in the context of the taxonomy's domain.

**Guideline 4:** *Demonstrate your taxonomy's applicability.* Especially when IS scholars conduct research on current topics and propose artifacts that are not yet available in practice, an evaluation of a taxonomy's applicability is feasible on the basis of illustrative scenarios (e.g., in an artificial environment [15]). They may find it useful to consult researchers working on a similar topic but not involved in building the taxonomy (e.g., [74] recruited other researchers), and may want to determine whether the taxonomy fulfills its (pre-defined) purpose. For example, if the taxonomy is built for classification, the representatives of the target group can be asked to classify objects and use the taxonomy accordingly. This allows to verify whether all objects can be classified and whether all the objects' characteristics are covered by a taxonomy. For example, [4] asks a potential user to classify objects by means of the developed taxonomy.

**Guideline 5:** *Evaluate your taxonomy's usefulness.* Especially when the taxonomy describes objects that are already used in practice, it is still possible to

**Figure 2. Frequency distribution of taxonomy evaluation criteria**



To the right is the long tail of taxonomy evaluation criteria with low frequency use, and to the left are few predominant criteria. Note that we nominalized criteria in case the original article formulated the criteria as verb.

evaluate the taxonomy's usefulness by means of a case study or field experiment (e.g., in a naturalistic environment [15]). To do this, practitioners may be provided with the taxonomy to evaluate its intended purpose (such as classification, differentiation, understanding, decision making, etc.). For example, if the taxonomy is built for decision-making (i.e., defining the what from guideline 1), representatives of the target group (i.e., defining for whom from guideline 1) can be asked to participate in a controlled experiment. The representatives are then assigned to either the treatment or the control group, both of which are introduced to the same use case. Unlike the control group, the treatment group additionally receives the taxonomy for facilitating decision making. Finally, *ceteris paribus*, the decision quality can be measured. For example, [47] evaluate their taxonomy's usefulness by means of a case study in which the taxonomy is used to guide the design of decision support systems.

**Guideline 6:** *Re-evaluate your taxonomy's robustness over a longer period of time.* There is no universal length of period recommended as it strongly depends on the volatility of the objects that the taxonomy describes. A very good opportunity would be a conference publication and a possible extended journal publication of a taxonomy research project. Naturally, there is a time in between both publications in which feedback on the revision of a taxonomy is usually obtained. If the number of objects increases or changes during this time period, this would provide an ideal opportunity to evaluate, for instance, the expressiveness and/or robustness of a taxonomy. For example, [77] developed an initial version of their taxonomy in 2017 [52], repeated the search for new objects in 2018 to see whether the taxonomy remains robust and provided a re-evaluated taxonomy in 2019.

## 6. Conclusion

Despite the great interest in taxonomies, current methodological guidance tends to focus on taxonomy building, leaving the need for guidance on evaluation. Ignoring concerns related to taxonomy evaluation can have significant negative consequences for such IS artifacts. Although there is a variety of DSR literature on evaluation, referring to taxonomy evaluation criteria, only very few are available to researchers for this type of artifact (e.g., [6] who specify criteria for certain types of artifact). In line with, for instance, Lukayneko et al. [76], artifact-specific guidelines for evaluation are a valuable starting point for planning and conducting appropriate evaluation. Drawing on our literature analysis, this study has several implications for research and practice: First, we provide a structured overview of the evaluation criteria currently applied to

taxonomies, which can be used by researchers and practitioners to select criteria in their projects. Surprisingly, as indicated by Figure 2, we obtained a variety of criteria from which about half are only employed in a single article, which is exposed by a 'long tail of taxonomy evaluation criteria'. Second, inspired by studying our sample of articles, we derive a set of six potential guidelines. These guidelines can be applied by, for instance, researchers and reviewers in testing and evaluating taxonomies. Third, with our work, we contribute to the body of knowledge related to DSR evaluation by providing the first set of artifact-specific evaluation criteria (e.g., for the entire class of 'constructs' March & Smith [6]), which can be extended and verified in further steps.

Although we derive helpful insights, this study is not free of limitations and opens avenues for future research. It is important to note that the frequency of taxonomy evaluation criteria employed in articles does not yet provide justificatory knowledge. This is why we plan to integrate our findings into further research. As a main direction, the preliminary set of guidelines and the taxonomy evaluation criteria require extensive evaluation itself to answer research questions such as - depending on the intended usage and evaluation context of a taxonomy - (1) which guidelines and evaluation criteria are best suited for which taxonomy, (2) the potential trade-offs of taxonomy evaluation criteria, (3) what measures to adopt when taxonomy evaluation fails, and (4) which evaluation methods are suitable for measuring taxonomy evaluation criteria (see [29] for more details).

As the establishment of guidelines presumes social interaction with the scientific community [76], the next steps should engage researchers in a dialogue to refine our set. With our study, we intend to take a very first step towards promoting the maturity of taxonomy evaluation. Thereby we seek to methodologically augment DSR in general and taxonomy research in particular.

**Acknowledgements.** This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Center "On-The-Fly Computing" (CRC 901, project number 160364472SFB901) and the European Regional Development Fund (ERDF) within the research project "SmartHybrid-Process Engineering" (ZW 6-85003451).

## 7. References

- [1] S. Gregor and A. R. Hevner, "Positioning and presenting design science research for maximum impact," *MIS Quarterly*, 37(2), 2013, pp. 337–356.
- [2] R. C. Nickerson, U. Varshney, and J. Muntermann, "A method for taxonomy development and its application in

information systems,” *European Journal of Information Systems*, 22(3), 2013, pp. 336–359.

[3] R. L. Glass and I. Vessey, “Contemporary application-domain taxonomies,” *IEEE Software*, 12(4), 1995, pp. 63–76.

[4] S. Gregor, “The nature of theory in information systems,” *MIS Quarterly*, 30(3), 2006, pp. 611–642.

[5] R. C. Nickerson, U. Varshney, and J. Muntermann, “Of taxonomies and taxonomic theories,” in *Proceedings of the 23rd Americas Conference on Information Systems*, Boston, MA, USA, 2017.

[6] S. T. March and G. F. Smith, “Design and natural science research on information technology,” *Decision Support Systems*, 15(4), 1995, pp. 251–266.

[7] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design Science in Information Systems Research,” *MIS Quarterly*, 28(1), 2004, pp. 75–105.

[8] B. Kuechler and V. Vaishnavi, “On theory development in design science research: Anatomy of a research project,” *European Journal of Information Systems*, 17(5), 2008, pp. 489–504.

[9] R. D. Galliers and F. F. Land, “Choosing appropriate information systems research approaches: A revised taxonomy,” in *Proceedings of the IFIP TC8 WG8.2*, 1990.

[10] M. Almufareh, D. Abaoud, and M. Moniruzzaman, “Taxonomy development for virtual reality (VR) technologies in healthcare sector,” in S. Chatterjee, K. Dutta and R. P. Sundarraj (eds.) *Designing for a Digital and Globalized World*, Springer, Boston, MA, USA, 2018, pp. 146–156.

[11] L. Püschel, M. Röglinger, and H. Schlott, “What’s in a smart thing? Development of a multi-layer taxonomy,” in *Proceedings of the 37th International Conference on Information Systems*, Dublin, Ireland, 2016.

[12] H. Gimpel, D. Rau, and M. Röglinger, “Understanding FinTech start-ups: A taxonomy of consumer-oriented service offerings,” *Electronic Markets*, 28(3), 2018, pp. 245–264.

[13] P. Haas, I. Blohm, and J. Leimeister, “An empirical taxonomy of crowdfunding intermediaries,” in *Proceedings of the 35th International Conference on Information Systems*, Auckland, New Zealand, 2014.

[14] K. Peffers, M. Rothenberger, T. Tuunanen, and R. Vaezi, “Design science research evaluation,” in *Proceedings of the 7th International Conference on Design Science Research in Information Systems and Technology*, Las Vegas, NV, USA 2012.

[15] C. Sonnenberg and J. vom Brocke, “Evaluations in the science of the artificial: Reconsidering the build-evaluate pattern in design science research,” in *Proceedings of the 7th International Conference on Design Science Research in Information Systems and Technology*, Las Vegas, NV, USA 2012.

[16] J. Venable, J. Pries-Heje, and R. Baskerville, “FEDS: A framework for evaluation in design science research,” *European Journal of Information Systems*, 25(1), 2016, pp. 77–89.

[17] F. Chasin, M. von Hoffen, M. Cramer, and M. Matzner, “Peer-to-peer sharing and collaborative consumption platforms: A taxonomy and a reproducible analysis,” *Information Systems and e-Business Management*, 16(2), 2018, pp. 293–325.

[18] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *Journal of Management Information Systems*, 24(3), 2007, pp. 45–77.

[19] Y. Wand, D. E. Monarchi, J. Parsons, and C. C. Woo, “Theoretical foundations for conceptual modelling in information systems development,” *Decision Support Systems*, 15(4), 1995, pp. 285–304.

[20] G. Lakoff, *Women, fire, and dangerous things. What categories reveal about the mind*. University of Chicago. Chicago, IL, USA, 1987.

[21] K. D. Bailey, *Typologies and taxonomies. An introduction to classification techniques*. Sage, Thousand Oaks, CA, USA, 1994.

[22] M. Schmidt-Kraepelin, S. Thiebes, M. C. Tran, and A. Sunyaev, “What’s in the game? Developing a taxonomy of gamification concepts for health apps,” in *Proceedings of the 51st Hawaii International Conference on System Sciences*, Waikoloa, HI, USA, 2018.

[23] D. H. Doty and W. H. Glick, “Typologies as a unique form of theory building: Toward improved understanding and modeling,” *Academy of Management Review*, 19(2), 1994, pp. 230–251.

[24] R. Bapna, P. Goes, A. Gupta, and Y. Jin, “User heterogeneity and its impact on electronic auction market design: An empirical exploration,” *MIS Quarterly*, 28(1), 2004, pp. 21–43.

[25] N. Prat, I. Comyn-Wattiau, and J. Akoka, “A taxonomy of evaluation methods for information systems artifacts,” *Journal of Management Information Systems*, 32(3), 2015, pp. 229–267.

[26] M. Holler, F. Uebernickel, and W. Brenner, “Defining archetypes of e-collaboration for product development in the automotive industry,” in *Proceedings of the 25th European Conference on Information Systems*, Guimarães, Portugal, 2017.

[27] M. Rosemann and I. Vessey, “Toward improving the relevance of information systems research to practice: The role of applicability checks,” *MIS Quarterly*, 32(1), 2008.

[28] T. G. Gill and A. R. Hevner, “A fitness-utility model for design science research,” in *Proceedings of the 6th International Conference on Design Science Research in Information Systems and Technology*, Milwaukee, WI, USA, 2011.



- [29] D. Szopinski, T. Schoormann, and D. Kundisch, "Because your taxonomy is worth it: Towards a framework for taxonomy evaluation," in *Proceedings of the 27th European Conference on Information Systems*, Stockholm-Uppsala, Sweden, 2019.
- [30] A. M. Oberländer, M. Röglinger, M. Rosemann, and A. Kees, "Conceptualizing business-to-thing interactions—A sociomaterial perspective on the Internet of Things," *European Journal of Information Systems*, 27(4), 2018, pp. 486–502.
- [31] M. M. Herterich, M. Holler, F. Uebernickel, and W. Brenner, "Understanding the business value: Towards a taxonomy of industrial use scenarios enabled by cyber-physical systems in the equipment manufacturing industry," in *Proceedings of the International Conference on Information Resources Management*, Ottawa, Canada, 2015.
- [32] S. Schneider, J. Lansing, F. Gao, and A. Sunyaev, "A taxonomic perspective on certification Schemes: Development of a taxonomy for cloud service certification criteria," in *Proceedings of the 47th Hawaii International Conference on System Sciences*, Waikoloa, HI, USA, 2014.
- [33] A. Strasser, "Delphi method variants in information systems research: Taxonomy development and application," *Electronic Journal of Business Research Methods*, 15(2), 2017, pp. 120–133.
- [34] R. Bärenfänger et al., "Classifying flexibility types in smart electric distribution grids: A taxonomy," in *Proceedings of the CIRED Workshop*, Helsinki, Finland, 2016.
- [35] T. Schäffer and D. Stelzer, "Towards a taxonomy for coordinating quality of master data in product information sharing", in *Proceedings of the 22nd MIT International Conference on Information Quality*, in *Proceedings of the 22nd MIT International Conference on Information Quality*, Little Rock, AR, USA, 2017.
- [36] D. Geiger, M. Rosemann, E. Fieft, and M. Schader, "Crowdsourcing information systems. Definition typology, and design," in *Proceedings of the 33rd International Conference on Information Systems*, Orlando, FL, USA, 2012.
- [37] T. Seyffarth, S. Kühnel, and S. Sackmann, "A taxonomy of compliance processes for business process compliance," in J. Carmona, G. Engels and A. Kumar (eds.) *Business Process Management Forum*, Springer, Berlin, Heidelberg, Germany, 2017, pp. 71–87.
- [38] N. M. Snow and J. L. Reck, "Developing a government reporting taxonomy," *Journal of Information Systems*, 30(2), 2016, pp. 49–81.
- [39] N. Fteimi and F. Lehner, "Analysing and classifying knowledge management publications. A proposed classification scheme," *Journal of Knowledge Management*, 22(7), 2018, pp. 1527–1554.
- [40] E. H. Diniz, E. S. Siqueira, and E. van Heck, "Taxonomy of digital community currency platforms," *Information Technology for Development*, 16(3), 2018, pp. 1–23.
- [41] A. Botha, M. Weiss, and M. Herselman, "Towards a taxonomy of mHealth," in *Proceedings of the International Conference on Advances in Big Data, Computing and Data Communication Systems*, Durban, South Africa, 2018.
- [42] P.H. Jarvinen, "Research questions guiding selection of an appropriate research method," in *Proceedings of the 8th European Conference in Information Systems*, Vienna, 2000.
- [43] E. Stoeckli, F. Uebernickel, and W. Brenner, "Capturing functional affordances of enterprise social software," in *Proceedings of the 23rd Americas Conference on Information Systems*, Boston, MA, USA, 2017.
- [44] U. Raza, W. Ahmad, and A. Khan, "Transformation from manufacturing process taxonomy to repair process taxonomy: a phenetic approach," *Journal of Industrial Engineering International*, 14(2), 2018, pp. 415–428.
- [45] A. Aksulu and M. Wade, "A comprehensive review and synthesis of open source research," *Journal of the Association for Information Systems*, 11(11), 2010, pp. 576–656.
- [46] M. Bock and M. Wiener, "Towards a taxonomy of digital business models. Conceptual dimensions and empirical illustrations," in *Proceedings of the 38th International Conference on Information Systems*, Seoul, Korea, 2017.
- [47] M. Siering, B. Clapham, O. Engel, and P. Gomber, "A taxonomy of financial market manipulations. Establishing trust and market integrity in the financialized economy through automated fraud detection," *Journal of Information Technology*, 32(3), 2017, pp. 251–269.
- [48] J. Zrenner, A. P. Hassan, B. Otto, and J. C. Marx Gómez, "Data source taxonomy for supply network structure visibility," in *Proceedings of the Hamburg International Conference of Logistics*, Hamburg, Germany, 2017.
- [49] W. R. King and V. Sethi Jr., "An empirical assessment of the organization of transnational information systems," *Journal of Management Information Systems*, 15(4), 1999, pp. 7–28.
- [50] K. Williams, S. Chatterjee, and M. Rossi, "Design of emerging digital services: a taxonomy," *European Journal of Information Systems*, 17(5), 2008, pp. 505–517.
- [51] F. H. A. Beevi, S. Wagner, S. Hallerstede, and C. F. Pedersen, "Data quality oriented taxonomy of ambient assisted living systems," in *Proceedings of the 1st IET International Conference on Technologies for Active and Assisted Living*, London, United Kingdom, 2015.
- [52] D. Szopinski, T. Schoormann, T. John, R. Knackstedt, and D. Kundisch, "How software can support innovating business models: A taxonomy of functions of business model development tools," in *Proceedings of the Americas Conference on Information Systems*, Boston, MA, USA, 2017.

- [53] A. Herzfeldt, M. Hausen, R. O. Briggs, and H. Krcmar, "Developing a risk management process and risk taxonomy for medium-sized IT solution providers," in *Proceedings of the European Conference in Information Systems*, Barcelona, Spain, 2012.
- [54] C. Gibbs, U. Gretzel, and J. Saltzman, "An experience-based taxonomy of branded hotel mobile application features," *Information Technology & Tourism*, 16(2), 2016, pp. 175–199.
- [55] S. Thiebes, G. Kleiber, and A. Sunyaev, "Cancer genomics research in the cloud. A taxonomy of genome data sets," in *Proceedings of the 4th International Workshop on Genome Privacy and Security*, Orlando, FL, USA, 2017.
- [56] S. Addas and A. Pinsonneault, "The many faces of information technology interruptions. A taxonomy and preliminary investigation of their performance effects," *Information Systems Journal*, 25(3), 2015, pp. 231–273.
- [57] J. Ge and U. Gretzel, "A taxonomy of value co-creation on Weibo. A communication perspective," *International Journal of Contemporary Hospitality Management*, 30(4), 2018, pp. 2075–2092.
- [58] G. Cledou, E. Estevez, and L. Soares Barbosa, "A taxonomy for planning and designing smart mobility services," *Government Information Quarterly*, 35(1), 2018, pp. 61–76.
- [59] R. Keller and C. König, "A reference model to support risk identification in cloud networks," in *Proceedings of the 35th International Conference on Information Systems*, Auckland, New Zealand, 2014.
- [60] K. Werder and H.-Y. Wang, "Towards a software product industry classification," in *Proceedings of the 15th International Conference on Intelligent Software Methodologies, Tools and Techniques*, Larnaca, Cyprus, 2016.
- [61] O. Labazova, T. Dehling, and A. Sunyaev, "From hype to reality: A taxonomy of blockchain applications," in *Proceedings of the 52th Hawaii International Conference on System Sciences*, Wailea, HI, USA, 2019.
- [62] S. Tönnissen and F. Teuteberg, "Towards a taxonomy for smart contracts," in *Proceedings of the 26th European Conference on Information Systems*, Portsmouth, United Kingdom, 2018.
- [63] D. E. Strode, "A dependency taxonomy for agile software development projects," *Information Systems Frontiers*, 18(1), 2016, pp. 23–46.
- [64] R. Tilly, O. Posegga, K. Fischbach, and D. Schoder, "Towards a conceptualization of data and information quality in social information systems," *Business & Information Systems Engineering*, 59(1), 2017, pp. 3–21.
- [65] M. Alrige and S. Chatterjee, "Toward a taxonomy of wearable technologies in healthcare," in *Proceedings of the 10th International Conference on Design Science Research in Information Systems and Technology*, Dublin, Ireland, 2015.
- [66] A. Gutierrez, J. Orozco, and A. Serrano, "Developing a taxonomy for the understanding of business and IT alignment paradigms and tools," in *Proceedings of the 16th European Conference on Information Systems*, Galway, Ireland, 2008.
- [67] K. Conboy, "Agility from first principles: reconstructing the concept of agility in information systems development," *Information Systems Research*, 20(3), 2009, pp. 329–354.
- [68] P. Tsatsou, S. Elaluf-Calderwood, and J. Liebenau, "Towards a taxonomy for regulatory issues in a digital business ecosystem in the EU," *Journal of Information Technology*, 25(3), 2010, pp. 288–307.
- [69] M. Fellmann, S. Robert, S. Büttner, H. Mucha, and C. Röcker, "Towards a framework for assistance systems to support work processes in smart factories," in *Proceedings of the IFIP International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, Reggio di Calabria, Italy, 2017.
- [70] F. Gao, S. Thiebes, and A. Sunyaev, "Rethinking the meaning of cloud computing for health care. A taxonomic perspective and future research directions," *Journal of Medical Internet Research*, 20(7), 2018, pp. 1–27.
- [71] J. Roeder, D. R. Cardona, M. Palmer, O. Werth, J. Muntermann, and M. H. Breitner, "Make or break. Business model determinants of FinTech venture success," in *Proceedings of the Multikonferenz Wirtschaftsinformatik*, Lüneburg, Germany, 2018.
- [72] J. Jöhnk, M. Röglinger, M. Thimmel, and N. Urbach, "How to implement agile IT setups. A taxonomy of design options," in *Proceedings of the 25th European Conference on Information Systems*, Guimarães, Portugal, 2017.
- [73] R. Barn and B. Barn, "An ontological representation of a taxonomy for cybercrime," in *Proceedings of the 24th European Conference on Information Systems*, Istanbul, Turkey, 2016.
- [74] K. R. T. Larsen, "A taxonomy of antecedents of information systems success. Variable analysis studies," *Journal of Management Information Systems*, 20(2), 2003, pp. 169–246.
- [75] T. Küpper, R. Jung, T. Lehmkuhl, and A. Wieneke, "Features for social CRM technology. An organizational perspective," in *Proceedings of the 20th Americas Conference on Information Systems*, Savannah, GA, USA, 2014.
- [76] R. Lukyanenko, J. Evermann, and J. Parsons, "Guidelines for establishing instantiation validity in IT artifacts: A survey of IS research," in *Proceedings of the 10th International Conference on Design Science Research in Information Systems and Technology*, Dublin, Ireland, 2015.
- [77] D. Szopinski, T. Schoormann, T. John, R. Knackstedt, and D. Kundisch, "Software tools for business model innovation: Current state and future challenges," to appear in *Electronic Markets*, 2019.